INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/acs.2739

Adaptive kernels in approximate filtering of state-space models

Kamil Dedecius*,†

Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou věží 1143/4, 182 08 Prague 8, Czech Republic

SUMMARY

Standard Bayesian algorithms used for online filtering of states of hidden Markov models from noisy measurements assume an accurate knowledge of the measurement model in the form of a conditional probability density function. However, this knowledge is often unreachable in practice, and the used models are more or less misspecified, or it is too complex, making the resulting models intractable. This paper focuses on these issues from the particle filtering perspective. It adopts the principles of the approximate Bayesian filtering, where the particle weights are based on the (dis)similarity of the true measurements and the pseudo-measurements obtained by plugging the state particles directly into the measurement equation. Specifically, a new robust method for online tuning of the weighting kernel is proposed. Copyright © 2016 John Wiley & Sons, Ltd.

Received 19 April 2016; Revised 19 August 2016; Accepted 23 September 2016

KEY WORDS: filtering; nonlinear filters; Bayesian filtering; sequential Monte Carlo, particle filter, approximate filtering

1. INTRODUCTION

Bayesian estimation heavily depends on a solid knowledge of measurements-generating models in the form of conditional probability distributions. However, to quote the famous statistician George Box, *"all models are wrong, but some are useful"* [1]. Most or all of the models describing real-world phenomena are more or less idealized, and do not completely reflect their intrinsic attributes, e.g., due to rarely occurring or not completely understood effects, or simply for mathematical convenience. The adopted models are thus incomplete and approximate, that is, to a certain degree *misspecified*. As a result, the true measurements may exhibit strong departures from them – the noise terms may have significantly different or completely unknown distributions, outliers may occur, some usually mild and neglected effects may be pronounced etc. The misspecification naturally impairs the parameter estimates, as the standard Bayesian procedures are generally not robust to it [2].

The present paper focuses on treating these issues in sequential estimation – filtering – of nonlinear state-space models, abounding, e.g., in signal processing, control, target tracking, econometrics, computer vision, statistics, biology, and many other fields [3–7]. It aims at the class of sequential Monte Carlo filters, called *particle filters*, and specifically at their approximate extensions.

The research in the particle filtering domain has devoted a significant attention to the misspecification issues. The solutions can be divided into several branches. One of them involves optimal (adaptive) distributions used for proposing new samples from the state space, taking the up-to-date measurements into account in order to stabilize and improve the sampling procedures [6, 8]. Although this approach is intuitively appealing, and also satisfies a sort of optimality criterion, it is

*Correspondence to: Kamil Dedecius, Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou věží 1143/4, 182 08 Prague 8, Czech Republic.
†E-mail: dedecius@utia.cas.cz

intractable in most cases [7]. The effort expended in attempting to approximate the optimal proposal distributions resulted in many methods, see, e.g., [6, 9]. An alternative way towards more robust particle filters consists in replacing the resampling procedure by a more sophisticated alternative, favouring particles that are more likely to survive at the next time step. This group of algorithms comprises primarily the auxiliary particle filters [10–12]. It is also possible to focus clearly on the case of outlier-contaminated measurements. The solutions involve second-order approximations of the log-likelihood [13], dynamic data rectification [14], and outlier detection via testing procedures [15]. A more general and stable approach is filtering with a bank of particle filters, approximating the nonnormal noise by normal mixtures [16]. Recently, an algorithm was proposed that deals with imprecise models even more directly by means of the random set theory framework [17, 18]. We note, that these basic algorithms gave rise to an abundance of various modifications, surveyed, e.g., in [4, 19–22]. This plethora of available algorithms, and the lack of a narrow *universal* class of solutions are often viewed as one of the main drawbacks of the particle framework.

A much more universal way around the discussed issues provide the novel approximate Bayesian filters, e.g. [23–25], allowing to approximate the posterior state distribution in cases where the measurement model is not sufficiently known or it is too complex to be evaluated. These filters, inspired by the approximate Bayesian computation methods (surveyed, e.g., in [26]), proceed with samples from the state space that are directly plugged into the measurement equation known, e.g., up to the noise properties. The resulting particle weights are proportional to the proximity of the obtained pseudo-measurements to the true measurement. The proximity is measured by means of a kernel function. The generic approximate filter adopts, similarly to the standard (offline) approximate Bayesian computation literature, a uniform kernel. It either accepts the particles and assigns them uniform weights, or rejects them if they yield pseudo-measurements too distant from the true measurement [24]. The recent results for the offline domain can be found, e.g., in [25–28], and in [29] proposing both online and offline (batch) algorithms allowing a gradient-based maximum likelihood estimation of static parameters in hidden Markov models with intractable likelihoods. It extends the earlier theoretical results [30] by using some ideas from [31].

The generic approximate filter has attractive asymptotic properties, ensuring convergence to a biased state estimator under fixed kernel and the number of particles going to infinity, with the bias tending to zero in the limit of a vanishing kernel scale [24]. However, this double convergence may be seen impractical [32], and a carefully designed algorithm for scale adaptation (contraction) is required to achieve reasonable convergence and prevent filter collapsing [33]. A conservative approach with good theoretical properties is the linear schedule, slowly contracting the scale in time [34]. An alternative approach determining the scale from the required effective sample size was reported in [35] and adopted in the generic approximate filter, too.

The uniform kernel used in the generic approximate filter has one more disadvantage: it completely neglects the proximity of the pseudo-measurements covered by the kernel to the true measurements, thus giving no information about the filter accuracy during the run time. It only assures acceptance or rejection of state samples, but the accepted ones are assigned with equal uniform weights. Although the filter is efficient in a long run, its short-time (or small sample) performance is impaired by this strategy. An alternative approach rooted in the kernel density estimation (KDE) was proposed recently [23]. It considers centered probability kernels with finite second-order moments. The scale adaptation exploits the standard KDE criterion minimizing the mean integrated square error, which inevitably calls for its subjective optimization due to the dependence of the optimal scale on the target density being estimated [36, Chap. 3.3–3.4]. With some additional simplifying assumptions, it results in a simple plug-in rule for scale determination. A potential drawback of the filter lies in the kernel construction: the kernel scales are independent of the respective values of the true measurements, and rely only on their stable evolution, well explained by the (*imprecise or intractable*) measurement model.

In this paper the problem of adaptive nonuniform kernels is studied from an alternative viewpoint. The idea is that if the true measurement model were known, its $p\%$ highest probability region (HPR) would cover a related set of possible measurements, including the pseudo-measurements. However, under the lack of its knowledge, it is surrogated by a convenient probability kernel providing this coverage. This principle allows to infer states that lead to pseudo-measurement covered by the HPR,

DOI: 10.1002/acs

and to ascribe the state particles weights proportional to the kernel values in a fashion of the standard particle filter.

The paper is organized as follows: Section 2 succinctly overviews basic principles of the Bayesian filtration of state-space models from the viewpoint of sequential Monte Carlo filtration and the approximate filtration. Section 3 develops the novel theory of adaptive kernels, discusses its properties and summarizes the result in Algorithm 1. Section 4 contains two simulation examples comparing the performance of the proposed solution with the state-of-the-art solutions and the particle filter under a well-defined and a misspecified model, respectively. Section 5 concludes the paper.

2. BACKGROUND ON BAYESIAN FILTERING

Let us consider the discrete-time state-space Markovian models with hidden state variables X_t determining measurements Y_t through

$$X_t | (X_{t-1} = x_{t-1}) \sim f(x_t|x_{t-1}), \tag{1}$$

$$Y_t | (X_t = x_t) \sim g(y_t|x_t), \tag{2}$$

where $t \in \mathbb{N}^0$ denotes the discrete time index, and $f(x_t|x_{t-1})$ and $g(y_t|x_t)$ are scalar or multivariate (stochastic) state and measurement functions, respectively. These functions may be linear or nonlinear. In Bayesian practice they are conditional probability densities, expressing the stochastic nature of the considered processes. The measurements y_t are directly observed, while the states x_t are hidden and inferred.

The Bayesian approach to sequential inference – filtering – of x_t from acquired measurements $y_{1:t} = \{y_1, \dots, y_t\}$ relies on a prior distribution with a probability density $\pi(x_0)$, quantifying the initial knowledge about the state variable based, e.g., on historical data or expert's opinion. This density, representing the Bayesian estimator (filter) of x_t, is sequentially updated by virtue of the Bayes' theorem

$$\pi(x_{0:t}|y_{1:t}) \propto \pi(x_0) \prod_{\tau=1}^{t} g(y_\tau|x_\tau) f(x_\tau|x_{\tau-1}). \tag{3}$$

That is, the state estimate is recursively evolved and updated according to (1) and (2), respectively.

The analytical tractability of the Bayesian update (3) is rare. One prominent case arises if $f(x_t|x_{t-1})$ and $g(y_t|x_t)$ are linear, or mildly nonlinear but locally differentiable (hence linearizable) functions with uncorrelated zero-mean normal stochastic components. Under normal prior density $\pi(x_0)$, the resulting filter is a Bayesian equivalent of the celebrated Kalman filter [37], or its extended variant [38]. Another tractable case arises if the space of X_t is discrete and finite. However, these conditions are rather restrictive in practice, as the functions may be nonlinear or their normality severely corrupted, e.g., by outliers. In such cases, the analytical approximations are not satisfactory and numerical integrations are problematic in high dimensions [39]. A feasible solution provide the sequential Monte Carlo methods.

2.1. Sequential Monte Carlo

The class of sequential Monte Carlo methods, also known as the particle filters [40], approximates the target density $\pi(x_{0:t}|y_{1:t})$ by a set of N Monte Carlo samples (termed particles) $x_t^{(i)}$ with weights $w_t^{(i)}$ in the point-mass sense

$$\pi(x_{0:t}|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta_{x_t^{(i)}}(x_t), \tag{4}$$

DOI: 10.1002/acs

where $\delta_{x_t^{(i)}}(x_t)$ is the Dirac delta measure located at $x_t^{(i)}$. The samples $x_t^{(i)}$ are obtained from a convenient proposal distribution $q\left(x_t^{(i)}|x_{0:t-1}^{(i)}, y_t\right)$. The weights $w_t^{(i)}$ taking values in the unit $(N-1)$-simplex are sequentially updated by the Bayes' theorem,

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{g\left(y_t \left| x_t^{(i)}\right.\right) f\left(x_t^{(i)} \left| x_{t-1}^{(i)}\right.\right)}{q\left(x_t^{(i)} \left| x_{0:t-1}^{(i)}, y_t\right.\right)}. \tag{5}$$

A significant research effort has been devoted to optimal design of the proposal distribution [6, 8]. However, for the sake of the present paper, we stick with the generic *bootstrap* filter [40], using the state function $f(x_t|x_{t-1})$ for this purpose. The propagation of samples through the state density provides a general, simple and powerful approach to filtering time series. The particle filter works well for standard problems where the model is a good approximation to the data [10]. The bootstrap variant of (5) thus simplifies to

$$w_t^{(i)} \propto w_{t-1}^{(i)} g\left(y_t | x_t^{(i)}\right). \tag{6}$$

A subsequent resampling mechanism is usually introduced to prevent the filter from collapsing due to a high concentration of weights mass to a few particles, or even to a single particle. Resampling constructs a new set of N particles by drawing with replacement from the original set $\left\{x_t^{(i)}\right\}_{i=1,\dots,N}$ with regard to $\left\{w_t^{(i)}\right\}_{i=1,\dots,N}$. A subsequent propagation through the proposal distribution spreads the particles in the regions of potentially high probability. Several resampling methods have been proposed, from the basic multinomial resampling, where the probability of $x_t^{(i)}$ being drawn is equal to its weight $w_t^{(i)}$, to systematic resampling methods minimizing the introduced variance [4, 39].

2.2. Approximate filters

The standard Bayesian filtering methods assume that the measurement model (2) is known in the form of a probability density. However, as it was discussed in Introduction, this (expressed) knowledge is often very limited due to the complexity of the observed process (typically the noise properties), or due to physical limitations. Alternatively, the model may be prohibitively complex and/or numerically intractable.

Assume, however, that there exists a form of the measurement model (2) that allows to simulate pseudo-measurements $u_t^{(i)}$ by plugging the state particles $x_t^{(i)}$ into it. This model may be a highly complex probability density, a differential equation (similarly to the static case studied in [41]), a stochastic process, or even a noise term-free equation. Then, the posterior density $\pi(x_{0:t}|y_{1:t})$ can be approximated by a density $\tilde{\pi}(x_{0:t}|y_{1:t})$ represented with samples $x_t^{(i)}$ from $f(x_t|x_{t-1})$, that produce pseudo-measurements $u_t^{(i)} \sim g\left(y_t|x_t^{(i)}\right)$ lying within a predefined neighborhood of the true y_t,

$$\begin{aligned}\tilde{\pi}(x_{0:t}|y_{1:t}) &= \pi(x_0) \int \tilde{\pi}(x_{1:t}, u_{1:t}|y_{1:t}) du_{1:t} \\ &\propto \pi(x_0) \prod_{\tau=1}^{t} \left[\int \tilde{g}_{y_\tau,\varepsilon_\tau}(u_\tau) g(u_\tau|x_\tau) du_\tau\right] f(x_\tau|x_{\tau-1}),\end{aligned} \tag{7}$$

where the kernel function $\tilde{g}_{y_\tau,\varepsilon_\tau}(u_\tau)$ defines the neighborhood. This is the generic approximate Bayesian filter [24]. Note, that the update (7) is an approximate variant of the Bayesian update (3), yielding the posterior importance weights $w_t^{(i)}$

$$w_t^{(i)} \propto w_{t-1}^{(i)} \tilde{g}_{y_t,\varepsilon_t}\left(u_t^{(i)}\right), \tag{8}$$

DOI: 10.1002/acs

c.f. Equation (6). Naturally, it is possible to benefit from sampling M pseudo-observations $\left\{u_t^{(i,j)}\right\}_{j=1,\ldots,M}$ for each particle $x_t^{(i)}$ and proceed with an averaged version of (8),

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{\sum_{j=1}^{M} \tilde{g}_{y_t,\varepsilon_t}\left(u_t^{(i,j)}\right)}{M},$$

as proposed in [35]. In any case, the choice of $\tilde{g}_{y_t,\varepsilon_t}(u_t)$ is a critical part of the filter design, and the present paper is devoted to this issue.

3. ADAPTIVE CHOICE OF $\tilde{G}_{Y_T,\varepsilon_T}(U_T)$

The generic approximate Bayesian filter [24], inspired by the static approximate Bayesian computation methods surveyed, e.g., in [26], constructs an admissible set of $u_t^{(i)}$ around the true measurement y_t with a radius ε_t,

$$\tilde{g}_{y_t,\varepsilon_t}\left(u_t^{(i)}\right) \propto \mathbb{1}_{\left[||u_t^{(i)}-y_t||\leqslant\varepsilon_t\right]}, \qquad \varepsilon_t > 0. \tag{9}$$

This $\tilde{g}_{y_t,\varepsilon_t}(\cdot)$ is also known as the uniform kernel [23]. In order to prevent the filter from collapsing, the scale (radius) ε_t must be properly tuned, e.g., to reach a preset effective sample size [35]. Although this approach enjoys good asymptotic properties, its uniform weighting strategy prevents the user from determination of more probable particles at particular time instants. This drawback is eliminated in [23], where the authors propose a procedure for online tuning of nonuniform kernels. It is based on the kernel density estimation (KDE) framework [36], and selects the optimal kernel scale based on the empirical distribution of $u_t^{(i)}$. However, only a suboptimal solution is reachable due to quite strong assumptions (e.g., the optimal scale depends on the unknown distribution being estimated). The procedure strongly relies on the past evolution of state particles, and neglects the location of the actual true measurement y_t, which can be seen as a drawback if the past data were not informative enough.

Finally, it should be remarked, that the kernel idea coincides with a noisy extension of the underlying hidden Markov process (1)–(2) with a new sequence of observations $\{Y_t + V_t\}$, where V_t is a random noise distributed according to the kernel $\tilde{g}_{y_t,\varepsilon_t}(\cdot)$. More on this interpretation can be found, e.g., in [29, Sec. 2].

In the subsequent section, a new method for evaluation of particle weights is devised, that stems from the generic approximate Bayesian filter, but to some degree exploits the attractive kernel-based viewpoint of the KDE-based filter.

3.1. Proposed kernel tuning procedure

The aim of the proposed procedure is to reflect the proximity of the simulated pseudo-measurements $u_t^{(i)}$ to the true measurement y_t under very mild assumptions. The underlying idea is that the $100p\%$ highest probability region (HPR) of the true measurements-generating model, i.e., the unknown or intractable conditional probability density $g(y_t|x_t)$, would cover the corresponding region of possible measurements, including the pseudo-measurements emerging from admissible particles $x_t^{(i)}$. However, the true model is unknown or intractable. Thus, we surrogate it by a convenient symmetric kernel $\tilde{g}_{y_t,\varepsilon_t}(u_t)$ centered at y_t, whose scale ε_t assures the coverage of a preset proportion of pseudo-measurements α/N by a p-HPR of this kernel. Indeed, this kernel is not equivalent to the true model, but ensures that the asymptotic convergence of the filter is preserved by conservation of a high proportion of admissible particles, favouring the more probable ones. The kernel must be invariant under a location change, and scalable, i.e., belong to the location-scale family of symmetric distributions [42]. Many popular distribution are members of this family, e.g., the normal, Cauchy, Student's t, exponential and Laplace distributions, to name a few.

DOI: 10.1002/acs

The kernel adaptation consists of two subsequent steps:

1. Finding $u_t^{[\alpha]}$, the pseudo-measurement that has the αth least distance from y_t. This step involves computing the distances $||y_t - u_t^{(i)}||$ for all $i = 1, \dots, N$.
2. Setting the kernel location parameter to y_t and the scale parameter ε_t so that $u_t^{[\alpha]}$ lies on the boundary of the resulting p-HPR, i.e.,

$$\left| \int_{y_t}^{u_t^{[\alpha]}} \tilde{g}_{y_t,\varepsilon_t}(u_t) du_t \right| = \frac{p}{2}. \tag{10}$$

This approach has several attractive features:

- The posterior particle weights are nonuniform and proportional to the proximity of the simulated pseudo-measurements $u_t^{(i)}$ to the true measurement y_t. This is particularly useful in the resampling step, where the more probable particles are replicated and those with negligible weights removed [43].
- The kernel construction around y_t is worthwhile if the particles have not sufficiently explored the state space yet. The particles are gradually concentrating around the true value of $x_t^{(i)}$ even in the case of their initial too high or too low dispersion.
- The method admits a wide class of kernels, including the heavy-tailed ones, better reflecting high-dispersion processes.
- Under most of the kernels, outlying y_t do not violate the filter stability, because the kernel scale becomes appropriately high, and the distant pseudo-measurements occupy relatively flat kernel tails. This will be demonstrated in Figure 2.
- The computation of the posterior weights (8) is principally the same as in the bootstrap particle filter, the only difference is that while the particle filter exploits the measurement model (likelihood), the approximate filter uses the kernel. The approximate filter is much more robust as it does not rely on a correct measurement model (2), of course at the cost of a slightly worse tracking performance if this knowledge is present.
- The computational complexity of the approximate filter depends on the used kernel. Compared with the bootstrap particle filter, the approximate filter with normal kernel increases the complexity only by (i) searching of $u_t^{[\alpha]}$, and (ii) tuning of ε_t; both these operations are very cheap with respect to the (identical) evaluation of posterior weights. On the other side, there are kernels (e.g. the Cauchy kernel) that provide a much cheaper computation of the posterior weights than the bootstrap particle filter assuming a normal measurement model.
- The asymptotic results of the generic particle filter are not influenced, i.e., the convergence of the filter is preserved.

3.2. *Computation of ε_t*

While the first step relies on standard sorting algorithms, the adaptation of ε_t is a slightly more complicated. Equation (10) requires that $u_t^{[\alpha]}$ is either a $(1-p)/2$-quantile of $\tilde{g}_{y_t,\varepsilon_t}(u_t)$ if $u_t^{[\alpha]} \leqslant y_t$, or $(1-p)/2 + p = (1+p)/2$-quantile if $u_t^{[\alpha]} \geqslant y_t$, respectively. Generally, the ρ-quantile is given by the quantile function. Denoting the cumulative distribution function of a random variable Ξ as

$$F(\xi) = \Pr(\Xi \leqslant \xi) = \rho, \qquad \rho \in [0, 1],$$

the quantile function is its inverse, $F^{-1}(\rho)$. If F belongs to the location-scale family, then the distribution of the variable $\Upsilon = a + \varepsilon\Xi$ with $a \in \mathbb{R}$ and $\varepsilon > 0$ belongs to this family too, and its quantile function has the form

$$Q(\rho) = a + \varepsilon F^{-1}(\rho), \qquad \rho \in [0, 1]. \tag{11}$$

With the help of (11) it is already possible to determine the scale parameter ε_t of $\tilde{g}_{y_t,\varepsilon_t}(u_t)$. We identify the location a with the true measurement y_t, and the ρ-quantile with $u_t^{[\alpha]}$. Instead of

DOI: 10.1002/acs

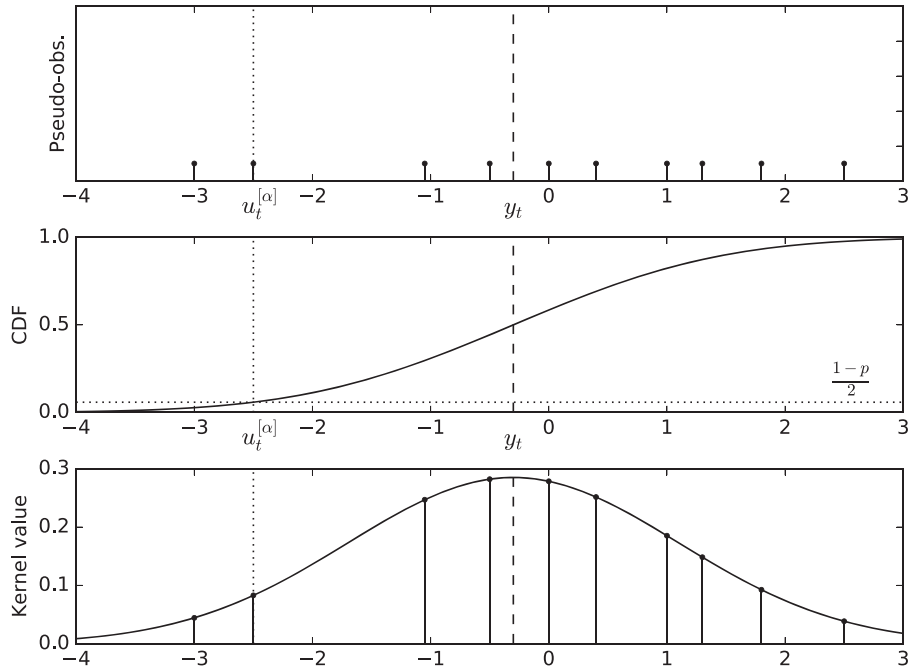

Figure 1. Scheme of the kernel tuning. Top: the ticks represent the pseudo-observations $u_t^{(i)}$, the dashed and dotted lines stand for the true observation y_t and the identified $u_t^{[\alpha]}$, respectively. Middle: The kernel scale parameter ε_t is sought such that $u_t^{[\alpha]}$ is a $(1-p)/2$-quantile of the adopted kernel. Alternatively, if $u_t^{[\alpha]}$ were greater than y_t, the symmetric $(1+p)/2$-quantile would be used. For better exposition, the cumulative distribution function (CDF) is used instead of its inverse – the quantile function. Bottom: The found scale ε_t parameterizes the kernel used for the determination of the (nonnormalized) particle weights.

determining whether $u_t^{[a]}$ is smaller or greater than y_t, it suffices to rely on kernel symmetry, and consider the nonnegative distance and the quantile right of y_t, i.e., $\rho = (1+p)/2$. Then,

$$\varepsilon_t = \frac{\left| Q\left(\frac{1+p}{2}\right) - y_t \right|}{F^{-1}\left(\frac{1+p}{2}\right)} = \frac{\left| u_t^{[\alpha]} - y_t \right|}{F^{-1}\left(\frac{1+p}{2}\right)}. \tag{12}$$

The principle of kernel scale tuning is depicted in Figure 1. The evaluation of quantile functions often requires numerical methods or approximations. However, in several important cases, closed-form expressions are available. Two prominent analytically tractable examples are given in the following section.

3.3. Examples of kernels

In this section, we derive the adaptation rules for the normal and Cauchy kernels, respectively. Figure 2 demonstrates the difference between these two kernels.

The normal kernel with a positive scale parameter (equivalently standard deviation) ε_t, translated to y_t, is given by

$$\tilde{g}_{y_t,\varepsilon_t}\left(u_t^{(i)}\right) \propto \exp\left(\frac{\left| u_t^{(i)} - y_t \right|^2}{\varepsilon_t^2}\right). \tag{13}$$

DOI: 10.1002/acs

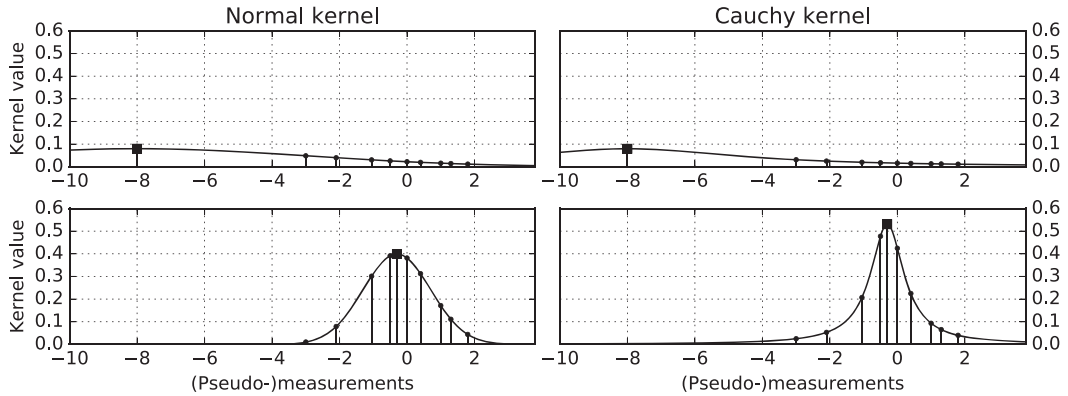

Figure 2. Normal and Cauchy kernel adaptation. The square marker denotes the true measurements y_t while the point markers represent pseudo-measurements $u_t^{(i)}$. If y_t is an outlier (top), the goal is to avoid degeneration of weights $w_t^{(i)}$ of $u_t^{(i)}$ by assignment of relatively noninformative (nearly uniform) weights via flat kernels. If y_t attains values close to $u_t^{(i)}$ (bottom), then the weights $w_t^{(i)}$ should appropriately reflect this closeness. The Cauchy kernel is narrower around the location parameter than the normal kernel, but has significantly heavier tails. The posterior weights are proportional to the value of the kernel.

The quantile function of $\mathcal{N}\left(y_t, \varepsilon_t^2\right)$ for $\rho \in [0, 1]$ has the form

$$Q(\rho) = y_t + \varepsilon_t \Phi^{-1}(\rho), \tag{14}$$

where $\Phi^{-1}(\cdot)$ denotes the quantile function of the standard normal distribution $\mathcal{N}(0, 1)$. That is, for $\rho = (1 + p)/2$ the rule for ε_t is given by

$$\varepsilon_t = \frac{\left|u_t^{([\alpha])} - y_t\right|}{\Phi^{-1}\left(\frac{1+p}{2}\right)}. \tag{15}$$

The Cauchy kernel is a heavy-tailed kernel with a positive scale ε_t, translated to y_t, and defined by

$$\tilde{g}_{y_t,\varepsilon_t}\left(u_t^{(i)}\right) \propto \left(1 + \frac{\left|u_t^{(i)} - y_t\right|^2}{\varepsilon_t^2}\right)^{-1}. \tag{16}$$

The quantile function of $\mathcal{C}\text{auchy}(y_t, \varepsilon_t)$ for $\rho \in [0, 1]$ has the form

$$\begin{aligned} Q(\rho) &= y_t + \varepsilon_t F^{-1}(\rho) \\ &= y_t + \varepsilon_t \tan\left[\pi\left(\rho - \frac{1}{2}\right)\right] \end{aligned} \tag{17}$$

where $F^{-1}(\cdot)$ denotes the quantile function of the standard Cauchy distribution $\mathcal{C}\text{auchy}(0, 1)$. For $\rho = (1 + p)/2$, the rule for ε_t is given by

$$\varepsilon_t = \frac{\left|u_t^{([\alpha])} - y_t\right|}{\tan\left(\frac{\pi p}{2}\right)}. \tag{18}$$

This kernel is particularly appealing for its heavy tails, suitable for filtration under outliers, and for its computational simplicity.

DOI: 10.1002/acs

Algorithm 1: Approximate filtration with adaptive kernels

Initialization:
Set the HPR value p, the number of covered pseudo-measurements α and the number of particles N.
Sample initial state particles $x_0^{(i)}, i = 1, \ldots, N$ from a suitable prior distribution $\pi_0(x_0)$.
Set uniform initial particle weights $w_0^{(i)} = 1/N$.
Online steps:
for $t = 1, 2, \ldots$ **do**
    Acquire true measurement y_t.
    Perform state prediction — Equation (1).
    Simulate pseudo-measurements $u_t^{(i)}$ by plugging $x_t^{(i)}$ into the measurement equation (2).
    Find $u_t^{[\alpha]}$, the αth least distant pseudo-measurement.
    Calculate kernel scale ε_t, Equation (12).
    Perform weights update, Equation (8).
    Resample.
end

4. EXAMPLES

The following two simulation examples demonstrate the robustness and tracking performance of the proposed kernel adaptation algorithm, and compares it with the state-of-the-art approaches, namely the generic method [24] and the KDE-based method with a plug-in rule [23]. In particular, the purpose of the first example is to show that the tracking performance is relatively close to the bootstrap filter employing a correct model. The second example demonstrates the superior properties of the proposed method under model misspecification. The mean square error (MSE) is used as the performance indicator,

$$\begin{aligned} MSE &= \frac{1}{t}\sum_{\tau=1}^{t}(\hat{x}_\tau - x_\tau)^2 \\ &= \frac{1}{t}\sum_{\tau=1}^{t}\left(\sum_{i=1}^{N} w_\tau^{(i)} x_\tau^{(i)} - x_\tau\right)^2. \end{aligned}$$

4.1. Example 1: Tractable model and comparison with the bootstrap particle filter

The first example is a benchmark proving that the performance of the proposed adaptive approximate filter is close to the performance of the bootstrap particle filter. We stress, that while the particle filter assumes full knowledge of the measurement model, the approximate filter alleviates this assumption and proceeds only with a noise term-free equation.

A simplified (linear) version of a state-space model popular in the particle filtering literature [28, 44] was adopted:

$$\begin{aligned} x_t &= \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos(1.2t) + v_t, \qquad && v_t \sim \mathcal{N}(0,1), \\ y_t &= x_t + w_t, && w_t \sim \mathcal{N}(0,10^2), \end{aligned} \tag{19}$$

initialized from $x_0 = 0$. The length of the series was 500 samples, the noise variables v_t and w_t were independent and identically distributed. Six variants of sequential Monte Carlo filters were compared: (A) the generic approximate filter [24] with a uniform kernel preserving 90% of samples, (B) the KDE-based approximate filter with a normal kernel [23], (C) the KDE-based filter with a quasi-Cauchy kernel [23], (D) the proposed approximate filter with a normal kernel, (E) the proposed approximate filter with a Cauchy kernel, and finally (F) the bootstrap particle filter with a full

DOI: 10.1002/acs

knowledge of the model (19). The same initial set of 1000 particles $\left\{x_0^{(i)}, i = 1, \ldots, 1000\right\}$ randomly sampled from the uniform distribution $\mathcal{U}(-100, 100)$ was used in all filters. The HPR value $p = 0.95$, and covered 90% of pseudo-measurements. A multinomial resampling was performed at each time step. 100 repetitions of the experiment were performed to obtain representative results.

The resulting statistics of the final mean square errors (MSE) are depicted in Figure 3 in the form of box plots, depicting the lower and upper quartile forming the box, the median MSE lying between them, and the whiskers whose length is 1.5 times the interquartile range. The particle filter (F) was very stable as expected, its final MSE values were very similar regardless of the noise realizations. The cases (D) and particularly (E) standing for the proposed method with adaptive normal and Cauchy kernels, respectively, are relatively close to the performance of the particle filter (F). The KDE-based filters were sensitive to the choice of kernel. While the quasi-Cauchy kernel (C) led to results close to the proposed method (D), the normal kernel (B) led to significantly worse tracking performance, similar to the generic method (A).

Figures 4 and 5 depict one particular realization of measurement noise w_t and the resulting settings of kernel scales, respectively. The character of evolution of ε_t is similar in both cases. The reactions to significant values of w_t will be more apparent in the next example.

Based on the simulation results it is possible to state that the performance of the proposed approximate filter is close to the performance of the bootstrap particle filter, which assumes full knowledge of the state-space model (19). Furthermore, the solution is not overly sensitive to the choice of kernel.

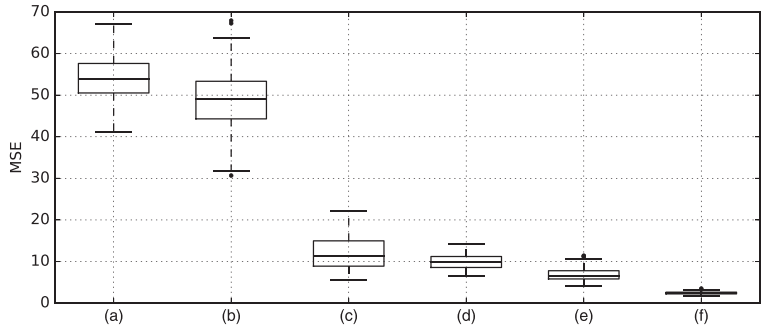

Figure 3. Example 1 — box plots of MSE of 100 independent experiment runs for (A) the generic approximate filter, (B) KDE-based filter with normal kernel, (C) KDE-based filter with quasi-Cauchy kernel, (D) proposed filter with normal kernel, (E) proposed filter with Cauchy kernel, and (F) the bootstrap particle filter. The dots depict outlying values.

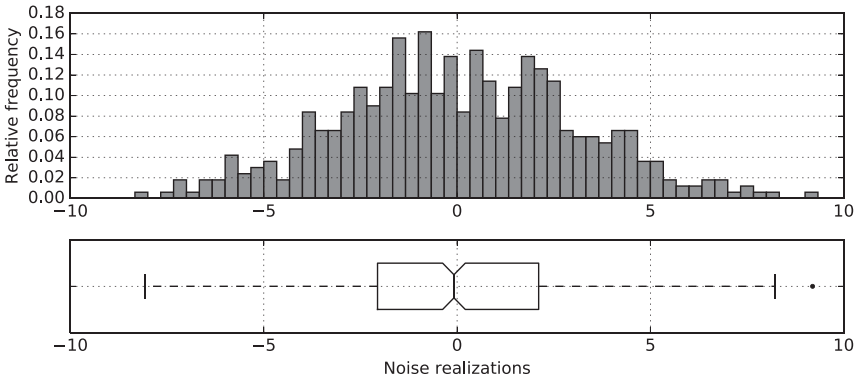

Figure 4. Example 1 — normal noise realization w_t. Histogram of relative frequencies and box plot. The box depicts upper and lower quartiles, between the lies the median. The whiskers have length of 1.5 times the interquartile range. The dot on the right-hand side depicts an outlier.

DOI: 10.1002/acs

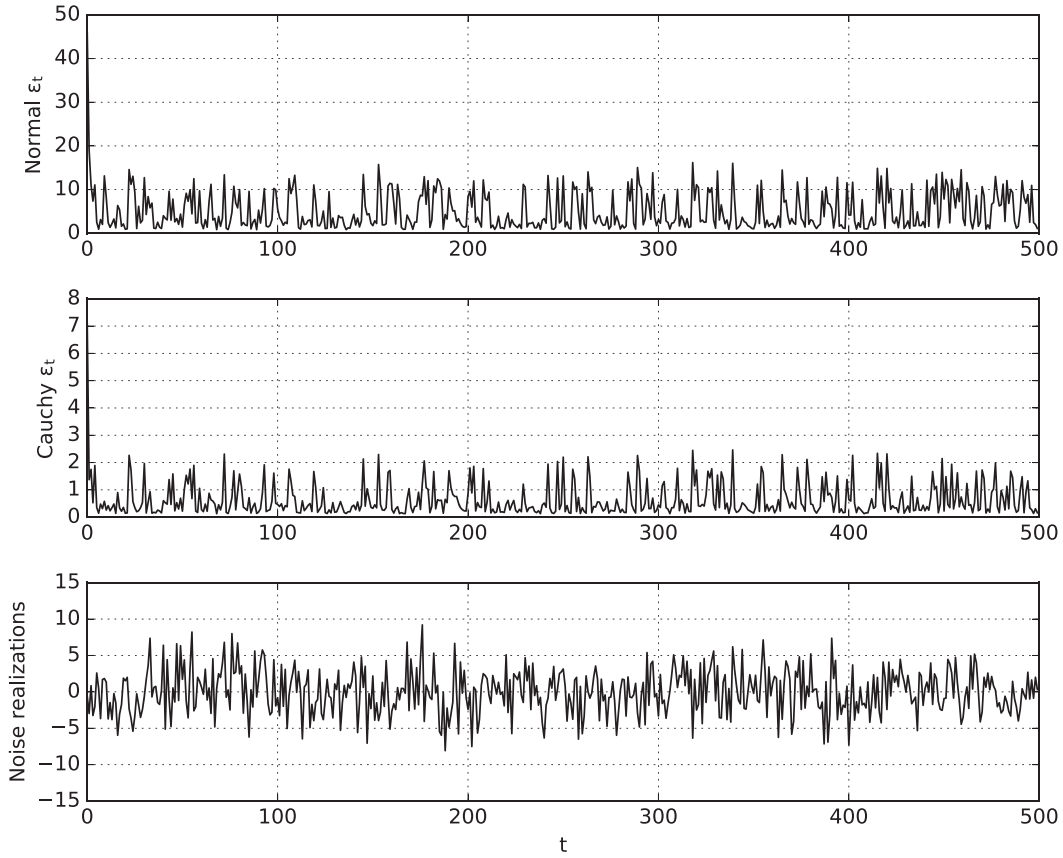

Figure 5. Example 1 — evolution of the normal and Cauchy scales ε_t (top, middle) with respect to normal noise realizations (bottom).

4.2. Example 2: Heavy-tailed stable noise

The purpose of the second example is to demonstrate the robustness of the proposed method in cases where the particle filter suffers tracking problems due to model misspecification, namely outliers. In particular, this example deals with observations corrupted by the Lévy's alpha-stable distributed noise [45]. More concretely, its special case – the Cauchy distribution – is used. It has undefined mean value and variance but unlike most other Lévy's alpha-stable distributions it has an analytically tractable probability density function. Nevertheless, the observation noise properties are ignored in the approximate filter, which allows its direct use in the other cases without any adjustments, while the standard particle filters would require more or less substantial modifications, often without a guarantee of success unless a tedious tuning is performed.

Let us consider the following model initialized from $x_0 = 0$,

$$\begin{aligned} x_t &= \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos(1.2t) + v_t, \qquad && v_t \sim \mathcal{N}(0,1), \\ y_t &= \frac{x_t^2}{20} + w_t, && w_t \sim \mathit{Cauchy}(0,1), \end{aligned} \tag{20}$$

and $t = 1, \ldots, 500$. This state-space model coincides with the popular nonlinear benchmark model frequently used in Monte Carlo literature (e.g., [28, 44]), with the exception for the measurement noise which, in our case, has the heavy-tailed Cauchy distribution located at the origin and with a unit scale. Figure 6 depicts an example of noise realizations used in one of the 100 independent experiment runs.

DOI: 10.1002/acs

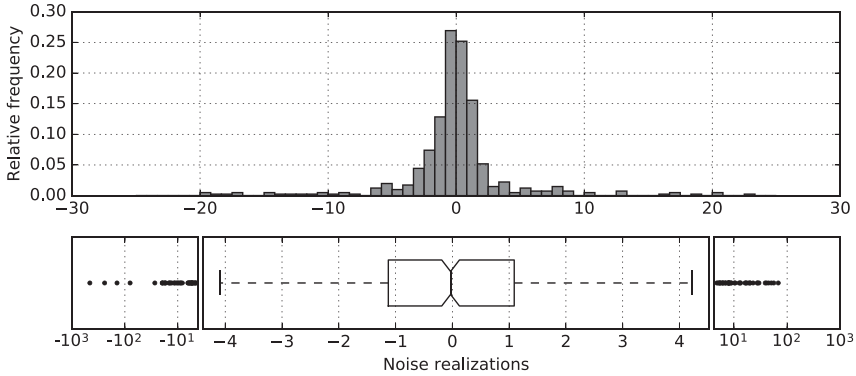

Figure 6. Example 2 — Cauchy noise realization w_t, histogram of relative frequencies limited to $[-30, 30]$, and scale-broken box plot (below). The box plot is divided intro three parts. The middle one depicts the box (upper and lower quartiles, and the median between them) and the whiskers of 1.5 times the interquartile range. Its scale is linear. The left and right parts depict the outliers and extreme values using logarithmic scales.

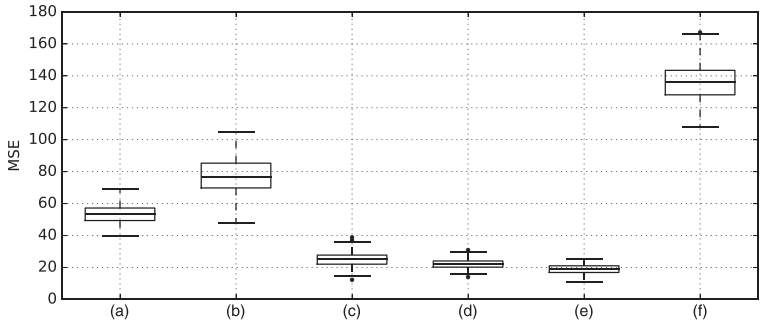

Figure 7. Example 2 — mean squared error box plots for 100 independent experiment runs for (A) the generic approximate Bayesian filter, (B) KDE-based filter with normal kernel, (C) KDE-based filter with quasi-Cauchy kernel, (D) proposed filter with normal kernel, (E) proposed filter with Cauchy kernel, and (F) the particle filter.

All six filters denoted again (A)—(F) started from the same initial set of 1000 particles sampled from $\mathcal{U}(-100, 100)$, the adaptive kernels of filters (D) and (E) were constructed to cover 90% of pseudo-measurements by a 95% HPR. A multinomial resampling procedure was performed each time step. The particle filter setting assumed a (misspecified) measurement model with a normal noise. This approach, standardly adopted in particle filtering, naturally leads to estimation difficulties, and calls for stabilization techniques discussed in Introduction. *The role of the particle filter in this example is only demonstrative and other filters are not compared with it*. On the other hand, the approximate filters plug the state particles $x_t^{(i)}$ directly into the *misspecified* variant of the measurement model (20). More precisely, an underspecified observation model without the noise term is assumed.

The box plots — Figure 7 — depict the MSE statistics. Again, it is apparent that the proposed filters are robust to the choice of kernels, with the Cauchy kernel providing slightly better tracking performance. Quite interestingly, the MSEs of all 100 runs of the proposed filters are very concentrated around the medians, which indicates high robustness to different heavy noise realizations. Finally, Figure 8 depicts the evolutions of normal and Cauchy kernel scales, and the noise time series in one experiment run. The adaptive method reflects well the noise evolution.

The conclusion of this example is that the proposed approximate filter is robust to the choice of kernel and has a superior stability under heavy-tailed noise, despite the underspecified measurement model.

DOI: 10.1002/acs

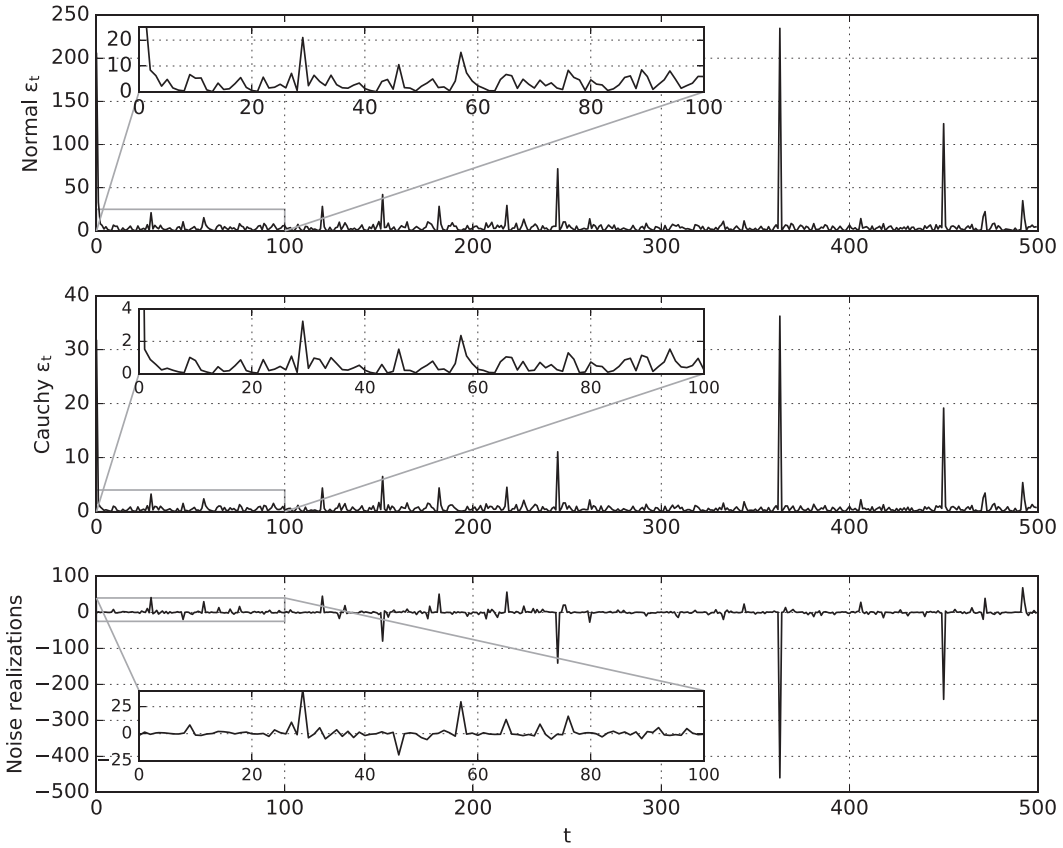

Figure 8. Example 2 — evolution of the normal and Cauchy scales ε_t (top, middle) with respect to Cauchy noise realizations w_t (bottom). Details of the first 100 values are given.

5. CONCLUSION

The class of approximate Bayesian filters provides excellent filtering properties in cases where the standard Bayesian methods suffer from a measurement model misspecification or intractability. This paper proposes an alternative approach to adaptation of the filter kernel, that is based on the assumption that the kernel should assure high coverage of the admissible pseudo-measurements generated by the filter. Two simulation examples demonstrate that the tracking performance is close to the bootstrap particle filter under well-specified model scenario, and under misspecification the proposed algorithm provides results superior to the state-of-the-art methods. In addition, the algorithm is not overly sensitive to kernel choice.

The future work involves an optimal setting of multivariate kernel under strong pseudo-measurements asymmetry, further acceleration possibilities (e.g., in a sense similar to [46]) and a focus on collaborative filtering in networks of cooperating agents studied, e.g., in [47, 48].

ACKNOWLEDGEMENTS

The work was supported by the Czech Science Foundation, project No. 14-06678P. The author is grateful to Dr. V. Sečkárová for helpful conversations and to two anonymous reviewers whose comments significantly helped to improve the paper.

REFERENCES

1. Box GEP. Robustness in the strategy of scientific model building. In *Robustness in Statistics*, Launer RL, Wilkinson GN (eds). Academic Press: New York, 1979; 201–236.
2. Miller JW, Dunson DB. *Robust Bayesian Inference via Coarsening*, 2015. arXiv:1506.06101 [stat.ME].

DOI: 10.1002/acs

3. West M, Harrison J. *Bayesian Forecasting and Dynamic Models* (Second ed.) Springer-Verlag: New York, 1997.
4. Doucet A, de Freitas N, Gordon N. *Sequential Monte Carlo Methods in Practice*. Springer: New York, 2001.
5. Simon D. *Optimal State Estimation, Kalman, H-Infinity, and Nonlinear Approaches*. Wiley-Interscience: New York, 2006.
6. Cappé O, Moulines E, Rydén T. *Inference In Hidden Markov Models*. Springer: New York, 2005.
7. Cappé O, Godsill SJ, Moulines E. An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE* 2007; **95**(5):899–924.
8. Doucet A, Johansen A. A tutorial on particle filtering and smoothing: Fifteen years later. In *Oxford Handbook of Nonlinear Filtering*, Crisan D, Rozovski B (eds). Oxford University Press: Oxford, 2009; 656–704.
9. Ristic B, Arulampalam S, Gordon NJ. *Beyond the Kalman filter: Particle filters for tracking applications*. Artech House: Norwood, USA, 2004.
10. Pitt MK, Shephard N. Filtering via Simulation: Auxiliary particle filters. *Journal of the American Statistical Association* 1999; **94**(446):590–599.
11. Johansen AM, Doucet A. A note on auxiliary particle filters. *Statistics & Probability Letters* 2008; **78**(12):1498–1504.
12. Pitt M, Silva R, Giordani P, Kohn R. *Auxiliary particle filtering within adaptive Metropolis-Hastings sampling*, 2010. arXiv:1006.1914 [stat.ME].
13. Li A, Jing Z, Hu S. Robust observation model for visual tracking in particle filter. *AEU-International Journal of Electronics and Communications* 2007; **61**(3):186–194.
14. Chen T, Morris J, Martin E. Dynamic data rectification using particle filters. *Computers & Chemical Engineering* 2008; **32**(3):451–462.
15. Maiz CS, Molanes-Lopez EM, Miguez J, Djurić PM. A particle filtering scheme for processing time series corrupted by outliers. *IEEE Transactions on Signal Processing* 2012; **60**(9):4611–4627.
16. Kotecha J, Djurić PM. Gaussian sum particle filtering. *IEEE Transactions on Signal Processing* 2003; **51**(10): 2602–2612.
17. Ristic B. Bayesian estimation with imprecise likelihoods: Random set approach. *IEEE Signal Processing Letters* 2011; **18**(7):395–398.
18. Mahler RPS. *Statistical Multisource-Multitarget Information Fusion*. Artech House Inc.: Norwood, USA, 2007.
19. Djurić PM, Kotecha J, Zhang J, Huang Y, Ghirmai T, Bugallo M, Miguez J. Particle Filtering. *IEEE Signal Processing Magazine* 2003; **20**(5):19–38.
20. Adali T, Haykin S. *Adaptive Signal processing: Next Generation Solutions*. Wiley-IEEE Press: New York, 2010.
21. Bruno MG. Sequential Monte Carlo Methods for Nonlinear Discrete-Time Filtering. *Synthesis Lectures on Signal Processing* 2013; **6**(1):1–99.
22. Van Leeuwen PJ, Cheng Y, Reich S. *Nonlinear Data Assimilation*. Springer: New York, 2015.
23. Calvet LE, Czellar V. Accurate methods for approximate Bayesian computation filtering. *Journal of Financial Econometrics* 2015; **13**(4):798–838.
24. Jasra A, Singh SS, Martin JS, Mccoy E. Filtering via Approximate Bayesian Computation. *Statistics and Computing* 2012; **22**(6):1223–1237.
25. Jasra A. Approximate Bayesian computation for a class of time series models. *International Statistical Review* 2015; **83**(3):405–435.
26. Green PJ, ŁAtuszyński K, Pereyra M, Robert CP. Bayesian computation: A perspective on the current state and sampling backwards and forwards. *Statistics and Computing* 2015; **25**(4):835–862.
27. Ehrlich E, Jasra A, Kantas N. Gradient free parameter estimation for hidden Markov models with intractable likelihoods. *Methodology and Computing in Applied Probability* 2015; **17**(2):315–349.
28. Martin JS, Jasra A, Singh SS, Whiteley N, Del Moral P, McCoy E. Approximate Bayesian computation for smoothing. *Stochastic Analysis and Applications* 2014; **32**(3):397–420.
29. Yildirim S, Singh SS, Dean T, Jasra A. Parameter estimation in hidden Markov models with intractable likelihoods using sequential Monte Carlo. *Journal of Computational and Graphical Statistics* 2015; **24**(3):846–865.
30. Dean TA, Singh SS, Jasra A, Peters GW. Parameter estimation for hidden Markov models with intractable likelihoods. *Scandinavian Journal of Statistics* 2014; **41**(4):970–987.
31. Poyiadjis G, Doucet A, Singh SS. Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika* 2011; **98**(1):65–80.
32. Marin JM, Pudlo P, Robert CP, Ryder RJ. Approximate Bayesian computational methods. *Statistics and Computing* 2012; **22**(6):1167–1180.
33. Biau G, Cèrou F, Guyader A. New Insights Into Approximate Bayesian Computation. *Annales de l'Institut Henri poincarè, Probabilitès et Statistiques* 2015; **51**(1):376–403.
34. Beskos A, Crisan D, Jasra A, Whiteley N. *Error Bounds and Normalizing Constants for Sequential Monte Carlo in High Dimensions*, 2011. arXiv:1112.1544 [stat.CO].
35. Del Moral P, Doucet A, Jasra A. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing* 2012; **22**(5):1009–1020.
36. Silverman B. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London, 1986.
37. Kalman RE. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 1960; **82**: 35–45.
38. Athans M, Wishner R, Bertolini A. Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements. *IEEE Transactions on Automatic Control* 1968; **13**(5):504–514.

DOI: 10.1002/acs

39. Kűnsch HR. Recursive Monte Carlo filters: Algorithms and theoretical analysis. *The Annals of Statistics* 2005; **33**(5):1983–2021.
40. Gordon N, Salmond D, Smith A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings of Radar and Signal Processing* 1993; **140**(2):107.
41. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface* 2008; **6**(31):187–202.
42. Lehmann EL, Casella G. *Theory of Point Estimation*. Springer: New York, 1998.
43. Bolić M, Djurić PM, Hong S. Resampling algorithms for particle filters: A computational complexity perspective. *EURASIP Journal on Advances in Signal Processing* 2004; **2004**(15):2267–2277.
44. Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2010; **72**(3):269–342.
45. Zolotarev VM. One-Dimensional Stable Distributions. In *Translations of mathematical monographs*, Vol. 65. American Mathematical Society: Providence, 1986.
46. Li T, Sun S, Corchado JM, Sattar TP, Si S. Numerical fitting-based likelihood calculation to speed up the particle filter. *International Journal of Adaptive Control and Signal Processing* 2015. (to appear).
47. Kárný M, Andrýsek J, Bodini A, Guy TV, Kracik J, Ruggeri F. How to exploit external model of data for parameter estimation? *International Journal of Adaptive Control and Signal Processing* 2006; **20**(1):41–50.
48. Dedecius K, Djurić PM. Diffusion filtration with approximate Bayesian computation. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; **2015**:3207–3211.

DOI: 10.1002/acs